EJLT European Journal of Law and Technology

The rise of the personal data quality principle: is it legal and does it have an impact on the right to rectification?

Diana Dimitrova*

Abstract

EU data protection law anchors the principle of data accuracy and connects it to the right to rectification. Both provisions have remained underexplored in literature, leaving a significant academic gap. The present paper will argue that in regulatory and academic discourse there is a tendency to replace 'data accuracy' with 'data quality.' The discourse seems to suggest that the concept of data quality subsumes and goes beyond classical accuracy elements, i.e., it is broader than accuracy. It seems to include requirements on the quality of data processing, not simply on the individual personal data pieces. Many of the non-exhaustive elements of data quality, as derived from non-legal disciplines, could be traced to different principles and legality requirements of EU data protection law, giving the concept of 'data quality' a legal foundation. As a result, the scope of the right to rectification could be broader than what it is currently argued to be and could give rise to a broader range of legitimate claims by data subjects.

Keywords: Algorithmic Profiling, Data Accuracy, Data Quality, Interdisciplinary Debate, Right to Rectification.

1. Introduction

The right to rectification in the EU data protection instruments grants data subjects the right to obtain from the controller the 'rectification of *inaccurate* personal data' which concerns him or her and, taking into account the purposes of the processing, the completion of *incomplete* personal data.[1] Thus, the main rationale behind the right to

* FIZ Karlsruhe – Leibniz Institute for Information Infrastructures and VUB/LSTS. The author wishes to thank Prof. Dr. Paul De Hert, Dr. Dara Hallinan, Prof. Dr. Gianclaudio Malgieri, Juraj Sajfert and the anonymous reviewers of this article for their precious comments on an earlier version of the article. The article is based on research performed in the framework of the author's PhD Dissertation, Diana Dimitrova, 'Data Subject Rights: The Rights to Access and Rectification in the Area of Freedom, Security and Justice', Vrije Universiteit Brussel, 2021.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ

rectification is argued to be the enforcement of the *principle of data accuracy and completeness* under data protection law.[2] The substance of the *legal* concept of personal data accuracy, however, remains an underexplored question, which is not clarified in the EU data protection framework. It is a crucial one, because the scope of the legal principle of data accuracy largely determines the scope of the right to rectification in data protection law and the respective obligations of the data controller when addressing such requests.

This paper sets out to explore the legal concept of (personal) data accuracy and completeness from a data protection point of view. It will demonstrate how data accuracy is becoming subsumed by the term 'data quality', which the paper will argue is emerging as an impactful term in EU data protection law. The paper proposes that data quality consists of the data protection principle of data accuracy and different quality requirements on the processing of the data, as will be discussed in the paper. The paper argues that the recognition of the principle of data quality would be an adequate response to the challenges of modern personal data processing technologies. The derived insights about *data quality* and *its relationship with data protection law*, especially with the data protection principles, will be used to explore the *scope* of the right to rectification. To build the argument, the paper is structured as follows.

First, it will briefly discuss the existing analytical literature on the principle of personal data accuracy as embedded in EU data protection law (Section 2). Second, the paper will present the ongoing academic and regulatory discussions on 'data quality' in the field of data protection and how these sometimes conflate it with the notion of 'data accuracy' (Section 3). Third, the paper will take stock of the rich body of interdisciplinary discussion on the concept of 'data quality'. It will focus especially on the work in relation to the development of the concept of data and/or information quality, deriving its dimensions and classifying them. It will also discuss the arguments in the literature that data quality imposes requirements not only those concerning individual data pieces, but also the different types of data processing operations, based on Liu and Chi's theory of the Data Evolution Life Cycle (DELC) (Section 4).

2016 L 119/1 (GDPR), Art 16; Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ 2016 L 119 (LED), Art 16 (1); Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ 2018 L 295 (Regulation 2018/1725), Art 18. Emphasis added.

[2] Tobias Herbst, 'Art. 16 Recht auf Berichtigung' in Jürgen Kühling and Benedikt Buchner (eds), *Datenschutz-Grundverordnung: Kommentar* (C.H.Beck 2017), 399, para 1; Enrico Peuker, 'Artikel 16 Recht auf Berichtigung' in Gernot Sydow (ed), *Europäische Datenschutzgrundverordnung: Handkommentar* (Nomos 2017), 478, para 27; GDPR, Art 5 (1) (d); LED, Art 4 (1) (d); Regulation 2018/1725, Art 4 (1) (d).

Fourth, the paper will discuss two examples of data quality issues related to the processing: algorithmic profiling and structural issues, especially from a data protection perspective (Sections 5 and 6). It will then discuss one more factor regarding the quality of data processing, namely human cognition (Section 7). With the discussion of data quality issues from a data protection perspective, the paper will address to some extent the criticism voiced by Hallinan and Borgesius that there is very little cross-fertilisation between the literature on the legal notion of accuracy and the interdisciplinary notion of data quality.[3] Finally, the paper will argue that if data quality is accepted as a legal notion which includes, but also goes beyond, the accuracy of personal data, then the right to rectification in EU data protection law has the potential to rectify not only factual mistakes in individual data pieces, but also problems of low quality related to processing operations. To this end, the paper will briefly study three examples from the point of view of data quality problems stemming from the quality of the *processing* operations and how the right to rectification might be applicable in these cases (Section 8). The paper will conclude with a summary of the concept of data quality and its influence on the scope of the right to rectification (Section 9).

Admittedly, the right to rectification might not be the only tool in data protection law which can address the identified data quality problems, because some of these problems might be also addressed by other rights, e.g., the right to erasure, restriction of processing and the right to object to the processing.[4] However, the present article will focus only on the right to rectification to demonstrate its potential and limits in addressing data quality problems in general.

2. Data protection in EU data protection law

Pursuant to the principle of data accuracy in EU law, the data controller should ensure that the personal data they process are 'accurate' and 'up to date'.[5] That principle links to the right to rectification, which provides data subjects with the possibility to request corrective actions in cases where the data are (1) not accurate or (2) not complete, depending on the purposes of the processing.[6] Although completeness is not mentioned explicitly in the definition of the principle of data accuracy, arguing that the concept of accuracy includes completeness would not be devoid of logic.[7]

However, it remains a fact that none of the EU data protection instruments define with sufficient precision the concepts of data accuracy, up-to-dateness and completeness. In addition, literature on the concept of data accuracy in data protection law, as well as case law on that topic, has been scarce.[8] This is problematic because in data protection law this

[3] Dara Hallinan and Frederik Zuiderveen Borgesius, 'Opinions can be incorrect! In our opinion: on data protection law's accuracy principle' [2020] 10 (1) International Data Privacy Law, see footnote 18.

[4] GDPR, Art 17, 18, 20 and LED, Art 16 (only on erasure and restriction of processing).

[5] GDPR, Art 5 (1) (d) in conjunction with the principle of accountability in Art 5 (2); LED and Regulation 2018/1725, Art 4 (1) (d).

[6] GDPR, Art 16; LED, Art 16 (1) and Regulation 2018/1725, Articles 18 and 82 (1).

[7] Herbst (n 2), 300, para 4.

[8] Historically speaking, the principle of data accuracy in data protection did not originate in the EU legal order, but has a longer history which dates back to Council of Europe data protection

has potentially significant consequences for data subjects, as it is not immediately clear in what respect personal data have to be inaccurate or incomplete in order to give rise to a justified claim for data rectification.[9]

So far, interpretative legal guidance on the concept of data accuracy has been provided by the Article 29 Working Party (now the European Data Protection Board (EDPB)). The Article 29 Working Party defines accuracy as data which are 'accurate as to a matter of fact'.[10] Inaccuracy of *factual data* relates to data which are not objectively accurate and do not correspond to reality.[11] Furthermore, law and case law have established that accuracy is purpose and context-dependent.[12] This means that data needs to be accurate enough for the specified purpose of the processing, 'but not more accurate than that'.[13] Thus, there is certain room for manoeuvre in the degree to which data have to be accurate.[14]

As to *non-factual data*, there is some debate as to whether the principle of accuracy applies also to non-factual data, such as inferences or opinions. There seems to be a general agreement that '*de jure*' the principle of data accuracy applies to non-factual data (i.e. inferential data) because they are still personal data. *A fortiori* the EU data protection instruments, which anchor the accuracy principle, apply to them too.[15] As noted by Hallinan and Borgesius, some authors *state* that the principle of data accuracy cannot *practically* apply to non-factual data, because it cannot be established whether such data are accurate or inaccurate. These, however, do not support their statement with arguments, similarly

---

instruments. Even then it was not defined by the lawmakers. For a historical overview of data accuracy and data quality in data instruments, see Jiahong Chen, 'The Dangers of Accuracy: Exploring the Other Side of the Data Quality Principle" [2018] 4(1) European Data Protection Law Review. See also Gloria Gonzalez Fuster, 'Inaccuracy as a privacy-enhancing tool" [2010] 12 Ethics and Information Technology, 87, 87-88; Hallinan and Borgesius (n 3).

[9] Herbst (n 2), 401, para 11.

[10] Article 29 Working Party, 'Guidelines on the implementation of the Court of Justice of the European Union judgment on "Google Spain and Inc v Agencia Espanola de Proteccion de Datos (AEPD) and Mario Costeja Gonzalez" C-131/12' (WP 225, 26 November 2014), 15.

[11] Herbst (n 2), 400, para 8 and 209, para 60.

[12] Article 5 (1) (d) GDPR. C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECLI:EU:C:2017:994, para 53.

[13] Gonzalez Fuster (n 8), 88. She also traces this purpose dependence back to the 1980 OECD Guidelines on the protection of privacy and transborder flows of personal data and to the Council of Europe Convention 108.

[14] This is somehow understandable since the GDPR applies to different sectors and thus should accommodate their specific needs and requirements. See Dara Hallinan and Frederik Zuiderveen Borgesius, 'GDPR Commentary: Article 5 GDPR: Principles relating to processing of personal data' in Franziska Boehm and Mark Cole (eds) *GDPR Commentary,* Elgar Publishers, forthcoming 2020. Franziska Boehm, *Information Sharing and Data Protection in the Area of Freedom, Security and Justice: Toward Harmonized Data Protection Principles for Information Exchange at EU-level* (Springer 2012), 11.

[15] Hallinan and Borgesius (n 3); Thomas Hoeren, 'Big data and the legal framework for data quality' [2017] International Journal of Law and Information Technology, 35; European Data Protection Supervisor, 'Guidelines on the Rights of Individuals with regard to the Processing of Personal Data' 25 February 2014, 18; Herbst (n 2), 430, para 8; Peuker (n 2), 474, para 7.

to those who *state* that accuracy can apply to inferential data.[16] By contrast, Hallinan and Borgesius convincingly *argue* that accuracy applies to inferential data.

To support their argument, they suggest that inferential data or opinions, which they use synonymously, consist of (1) factual input data, which are processed on the basis of (2) an interpretative framework. They agree with the existing literature that the accuracy principle undoubtedly applies to factual data. They further argue that it also applies to the interpretative framework, when this framework is based on some (scientifically) recognised professional standard, e.g., about medical diagnoses, about decisions made when applying a certain legal framework to the situation of an individual.[17] Their conclusion is logical bearing in mind the fact that EU data protection law seeks to protect not simply personal data in the abstract. It seeks to protect individuals in relation to the *processing* of their personal data,[18] including when the processing consists in complex analytical and predictive analyses, as compared to the mere storage of the data. Because the risks largely depend on the processing, then what is (in)accurate can be determined on a case-by-case basis in light of the applicable processing framework. However, it should be noted that it is questionable whether the processing framework, e.g., an algorithm, can always be of scientifically proven quality. A prime example is the establishment of risk profiling algorithmic criteria based on indicators in the law enforcement and migration field, which are supposed to be set up by regulators and executive officers.[19]

Judicial proceedings in criminal law constitute a special exception to the accuracy principle. For example, the Law Enforcement Directive (LED) clarifies that subjective statements of individuals cannot always be verified.[20] In those cases the principle of accuracy, examined in light of 'the nature and purpose of the processing', should be understood to mean only the fact that a statement has been made and should not concern the content of the (witness, victim, suspect) statement.[21] The present paper notes that this should be without prejudice to the fact that the *facts* in such statements could be verified, whereas it could be said that the 'subjective perceptions'[22] of the person making a statement cannot be verified. This is presumably because there seem to be no accuracy standards which pertain to personal, subjective thoughts and impressions about the world and people around us.

A further aspect of the academic and regulatory discussion on the concept of data accuracy in EU data protection law is that it sometimes relates to the concept of 'data quality.' This

---

[16] Hallinan and Borgesius (n 3).

[17] Ibid.

[18] Lorenzo Dalla Corte, 'Scoping Personal Data: Towards a Nuanced Interpretation of the Material Scope of EU Data Protection Law' [2019] 10 (1) *EJLT*, https://ejlt.org/index.php/ejlt/article/view/672.

[19] Regulation (EU) 2018/1240 of the European Parliament and of the Council of 12 September 2018 establishing a European Travel Information and Authorisation System (ETIAS) and amending Regulations (EU) No 1077/2011, (EU) No 515/2014, (EU) 2016/399, (EU) 2016/1624 and (EU) 2017/2226 [2018] OJ L 236/1 (ETIAS Regulation), art 33 (3), (4) and (6).

[20] E.g. see LED, recital 30: 'The principle of accuracy of data should be applied while taking account of the nature and purpose of the processing concerned.'

[21] Ibid.

[22] Ibid.

raises the question of what the concept of data quality is and what its relationship with data accuracy is, which will be studied in the next section. It is especially interesting to explore whether it differs from the scope of data accuracy and if so, what the consequences for the rights of data subjects, and in particular, the right to data rectification are.

3. Data accuracy as an element of data quality?

Similar to data accuracy, 'data quality' is not defined in the EU data protection laws. As Schafer has noted, '(t)o the extent that it is studied at all, the interaction between law and information quality is typically seen as a relatively recent development promoted by the desire to regulate the collection, dissemination and use of ever larger amounts of data by public authorities.'[23] This is problematic not only because it is not clear what is meant by data quality, but also because the term 'data quality' is increasingly beginning to be used interchangeably with 'data accuracy',[24] indicating a possible overlap between both terms. For example, recently the Article 29 Working Party argued that keeping data accurate enhances their quality.[25] Further, in the General Data Protection Regulation (GDPR) in the context of Binding Corporate Rules (BCR), 'data quality' is mentioned in a context which suggests that it refers to data accuracy as enshrined in Article 5 (1) (d) GDPR.[26] More precisely, the said BCR provision mentions the applicability of the data protection principles in Article 5 GDPR to BCRs and explicitly names these principles as they are termed in Article 5 GDPR. There is one significant exception: rather than putting 'data accuracy', the legislator chose 'data quality' instead. This gives rise to two hypotheses: (1) that the concept of data quality might be seen as a type of legal and data protection concept, and

---

[23] Burkhard Schafer, 'Information Quality and Evidence Law: A New Role for Social Media, Digital Publishing and Copyright Law?' in Luciano Floridi and Phyllis Illari (eds), *The Philosophy of Information Quality* (Springer, 2014), 217.

[24] The term 'data quality' was used earlier in data protection literature, e.g. in Article 29 Working Party opinions, to refer to four of the core principles of data protection – purpose limitation, data accuracy, data minimisation and limited storage (See Article 29 Working Party, 'Guidelines on consent under Regulation 2016/679' (WP259 rev.01, 10 April 2018); Article 29 Working Party, 'Opinion on some key issues of the Law Enforcement Directive (EU 2016/680)' (WP 258, 29 November 2017); Article 29 Working Party, 'Opinion 03/2013 on purpose limitation' (WP 203, 02 April 2013); Article 29 Working Party, 'Opinion 03/2010 on the principle of accountability' (WP 173, 13 July 2010); Article 29 Working Party, 'Working Document on Black Lists' (WP 65, 3 October 2002); Article 29 Working Party, 'Opinion 07/2007 on data protection issues related to the Internal Market Information System (IMI)' (WP 140, 20 September 2007), 8 and 11. Under the earlier EU data protection framework, namely the repealed Directive 95/46/EC and repealed Regulation 45/2001, these principles were framed as 'principles relating to data quality,' but the reference to data quality was removed from the current EU data protection framework (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the pro-cessing of personal data and on the free movement of such data, OJ 1995 L 281/31, Art 6 and Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data, OJ 2001 L 8/1, Art 6). See by contrast GDPR, Art 5; LED, Art 4; Regulation 2018/1725, Art 4.

[25] Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (WP251rev.01, 6 February 2018), 11 -12.

[26] GDPR, Art 47 (2) (d).

(2) that data quality might be seen as an alternative to, or the same as, the concept of data accuracy. This is slightly different from the Article 29 Working Party formulation, which suggests that accuracy might be a function of data quality, but not the same as data quality (hypothesis (3)).

Particularly hypotheses (1) and (3) seem to find support especially in two (other) legal sources. First, come the EU data protection provisions applicable in the law enforcement context. These require the Member State competent law enforcement authorities to verify the *quality* of data before transmitting them to other competent authorities by providing 'necessary information' on the 'degree of accuracy, completeness and reliability of personal data, and the extent to which they are up to date.'[27] This provision suggests that the term 'data quality' might not be simply synonymous with data accuracy, but that it might be an umbrella term which includes accuracy and up-to-dateness, but also further requirements or concepts, such as *reliability*. The applicable data protection provisions in the law enforcement field anchor two further quality requirements: to distinguish between (a) facts and opinions and (b) data subjects, e.g. suspects, victims, witnesses.[28] Such categorisation of the data subjects is embedded, for example, in the structure of the Schengen Information System, which distinguishes between alerts in relation to suspects, victims, entry bans, witnesses, etc.[29] In short, the quoted examples demonstrate that for an accurate processing of personal data it is not enough that the personal data be factually accurate. They need to fulfil also a wide range of quality requirements, e.g., on data structuring.

Second, the EU AI Act Proposal explicitly refers both to accuracy and quality as safeguards in the framework of high-risk AI. What is interesting is that in the Proposal *quality* seems

[27] LED, art 7 (2) and Regulation 2018/1725, art 74.

[28] LED, art 6 and 7 (1). These principles were also recommended by the EDPS, See European Data Protection Supervisor, 'Opinion of the European Data Protection Supervisor on the Proposal for a Council Framework Decision on the exchange of information under the principle of availability (COM (2005) 490 final)' OJ C 116, 17 May 2006, para IV.6).

[29] SIS II (still applicable): Regulation (EC) No 1987/20061 of the European Parliament and of the Council on the establishment, operation and use of the second generation Schengen Information System (SIS II) [2006] OJ L381/4; Council Decision 2007/533/JHA on the establishment, operation and use of the second generation Schengen Information System (SIS II) [2007] OJ L205/63. SIS III (not yet applicable): Regulation (EU) 2018/1862 of the European Parliament and of the Council of 28 November 2018 on the establishment, operation and use of the Schengen Information System (SIS) in the field of police cooperation and judicial cooperation in criminal matters, amending and repealing Council Decision 2007/533/JHA, and repealing Regulation (EC) No 1986/2006 of the European Parliament and of the Council and Commission Decision 2010/261/EU [2018] OJ L312/56; Regulation (EU) 2018/1861 of the European Parliament and of the Council of 28 November 2018 on the establishment, operation and use of the Schengen Information System (SIS) in the field of border checks, and amending the Convention implementing the Schengen Agreement, and amending and repealing Regulation (EC) No 1987/2006 [2018] OJ L312/14; Regulation (EU) 2018/1860 of the European Parliament and of the Council of 28 November 2018 on the use of the Schengen Information System for the return of illegally staying third-country nationals [2018] OJ L312/1.

to be a requirement for the training, validation, and testing data sets for AI technologies.[30] These have to fulfil a non-exhaustive list of criteria, amongst which those in relation to the design choices, modelling, and the assumptions which the models are supposed to represent. In addition, they have to be sensitive to the context of the future usage of the AI technology, and be free from bias, amongst other things.[31] The Proposal also refers to the compliance with these requirements as necessary for the operation of high-risk AI technologies, even where their development does not involve the usage of training data.[32] At the same time, the Proposal also refers to accuracy, especially as a requirement to be fulfilled when high-risk AI is operational, and whose level and metrics have to be indicated to the users of the AI technologies.[33] This parallel mention of both terms does not specify their relationship and does not bring clarity about how they exactly differ from each other. It is interesting, though, that the mention of 'quality' in relation to design choices, to the assumptions about the problem to be solved via AI technology, to the context of deployment of the technology, and to the proper modelling of the data, could signal that these are not part of the accuracy concept, in slight contrast to Hallinan and Borgesius's proposition that accuracy applies also to the processing framework. It seems clear, though, that it adds to the accurate performance of high-risk AI.

Briefly put, it looks like personal data quality is a multifaceted notion which is featuring more prominently in legal instruments. The following Sections will explore in more detail the interdisciplinary literature on the different dimensions and elements encompassed by data quality and propose how they should be understood to relate to data accuracy, including other principles in EU data protection law. The exploration starts with an overview of the debate on data quality in other disciplines.

4. Data quality is purpose-dependent and multi-faceted

The call for attention on data quality started in the 1960s in the framework of the then, emerging computerised information.[34] Despite this call for attention to data quality, Bygrave signalled in the 1990's that whereas the development of (computing) technologies for data processing had been advancing fast in the preceding decades, the issue of data quality had remained unaddressed.[35]

---

[30] European Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (COM) 2021 206 final, Brussels, 21 April 2021 ('Proposed EU AI Act'), recital 38, 43, 44 and Art 10 (1) to (5).

[31] Proposed EU AI Act, Art 10 (1) to (5).

[32] Proposed EU AI Act, Art 10 (6).

[33] Proposed EU AI Act, recital 33, 49 and Arts 13(3) and 15.

[34] Kenneth L. Karst, ''The Files': Legal controls over the accuracy and accessibility of stored personal data' [1966] 31 (2) (4) Law and Contemporary Problems, 343.

[35] Lee A. Bygrave, 'Ensuring Right Information on the Right Person(s), Legal Controls of the Quality of Personal Information, Part I,' [1996], <https://www.jus.uio.no/ifp/om/organisasjon/afin/forskning/notatserien/1996/4_96.html> accessed 04 May 2019.

In the computer science field, this has been changing since the 1990s. An instructive summary of the work on information quality performed in the last decades has been undertaken by Illari and Floridi in their book on Information Quality of 2014. [36] They observe that while literature and practice on information quality has grown incrementally, it is challenging to offer a comprehensive overview of all the wok in the field. However, they have presented a concise overview of the interdisciplinary work on information quality.[37] Thus, the following paragraphs cannot strive for exhaustiveness. Instead, they will focus on the following two aspects: (1) identifying and categorising the different data quality dimensions. The discussion will also include Liu and Chi's criticism on the lack of a theoretical framework for examining data quality and their proposed solution (2).

Illari and Floridi note that work on information quality originated in the computer science field at the Massachusetts Institute of Technology (MIT) in the 1990s.[38] MIT scientists, e.g., Wand, Wang and Strong, focused their studies especially on the information quality needs of data consumers in the business management context.[39] Batini and Scannapieco have also done substantial research on data quality dimensions and models, amongst others.[40]

On a substantive level, MIT scientists argue that good quality information is information which is '*fit for purpose*, going far *beyond mere accuracy* of information'.[41] Similarly, Bygrave defines information quality as 'various characteristics or attributes of information which bear on the worth of the latter for *given purposes* and *given persons*'.[42] This is because the different users have their own requirements and define data quality according to their needs and the purposes of the processing.[43] This finding concerning the purpose-dependence of data quality demonstrates its similarity with the legal concept of data accuracy, as discussed in Section 2 above. Furthermore, scientists have argued that the accuracy of a certain piece of data does not guarantee the good quality of the data and their processing, because factors such as the format and structure of the data might make

[36] Luciano Floridi and Phyllis Illari (eds), *The Philosophy of Information Quality* (Springer, 2014).
[37] Ibid, 12.
[38] Ibid, 5.
[39] Ibid, 6.
[40] Carlo Batini and Monica Scannapieco, *Data and Information Quality: Dimensions, Principles and Techniques* (Springer 2016). Carlo Batini and Monica Scannapieco, *Data Quality: Concepts, Methodologies and Techniques* (Springer 2006).
[41] Floridi and Illari (n 36), 6. They refer in particular to the work of Wang, Wand and Strong. Emphasis added. Similar understanding suggested in European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Data Protection Law* [2018], 127.
[42] Lee A. Bygrave, *Data Protection Law: Approaching Its Rationale, Logic and Limits* (Wolters Kluwer International, Information Law Series 2002), 25, Emphasis added.
[43] Richard Y. Wang, M.P. Reddy and Amar Gupta, 'An object-oriented implementation of quality data products' [1993] Paper presented at the WITS, Orlando.

it less fit for the purposes of a certain processing.[44] Therefore, they argue, that absolute, 'intrinsic' or 'inherent' data quality does not exist.[45]

In addition, the finding that quality goes beyond accuracy confirms the observations derived from the legal discussion in Section 3 on data quality, namely that quality is a concept broader than accuracy, but which seeks to ensure the accurate processing of the data and is echoed in discussions, e.g., on Big Data. For example, one of the core attributes of Big Data is 'veracity', which refers to data quality as accuracy and overall quality[46] and 'documenting quality and uncertainty'.[47] Sometimes data quality is referred to indirectly, by referring to low data quality,[48] which is framed as data which are poorly selected/unrepresentative, incomplete, outdated and incorrect.[49] This means that accuracy might not be the only factor ensuring the quality of the processing. Other factors are also significant and all of them together ensure the accuracy of the data processing and its results.

Mentioning the additional criteria for data quality, beyond accuracy, raises the question what all the different elements of data quality are, i.e., if they can be all exhaustively listed. Scholars outside the legal field have performed impressive work on *identifying and categorising* the dimensions of information quality.[50] This work, to which many scientists have contributed, has been summarised by Lee et al in a table, as re-printed by Illari and

---

[44] Suzanne M. Embury and Paolo Missier, 'Forget Dimensions: Define Your Information Quality Using Quality View Patterns' in Floridi and Illari (n 36), 26; Wang and Strong, 'Beyond Accuracy: What Data Quality Means to Data Consumers,' Journal of Management Information Systems, Vol. 12, No. 4 (Spring, 1996), 5-33, 6 and 9.

[45] Phyllis Illari, 'IQ: Purpose and Dimensions' in Floridi and Illari (n 36), 284 – 285 and 300.

[46] Nikolaus Forgó, Stefanie Hänold and Benjamin Schütze 'The Principle of Purpose Limitation and Big Data' in M. Corrales et al. (eds.), *New Technology, Big Data and the Law* (Springer Nature Singapore, Pte Ltd 2017), 21 and 26. They argue it includes trustworthiness, accessibility, and being fit for the purpose of the processing. At the same time these authors indirectly refer to data accuracy as data quality when they enumerate the main data protection principles in Article 6 of the abolished Directive 95/46/EC. See also Mario Callegaro and Yongwei Yang, 'The Role of Surveys in the Era of "Big Data"' in David L Vannette and Jon A. Krosnick (eds.), *The Palgrave Handbook of Survey Research* (Palgrave Macmillan 2018), 175-176.

[47] Edward Curry, 'The big data value chain: definitions, concepts, and theoretical approaches' in Jose Maria Cavanillas, Edward Curry, Wolfgang Wahlster W (eds) *New horizons for a data-driven economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (Springer 2016), 30.

[48] Ibid, 21.

[49] European Union Agency for Fundamental Rights, '#BigData: Discrimination in data-supported decision-making' FRA Focus, 29 May 2018, 5. In that respect, FRA refers to 'The White House (2016), Big Data: A report on Algorithmic Systems, Opportunity and Civil Rights.'

[50] Floridi and Illari (n 36), 7-8. Floridi and Illari summarised the different approaches to identifying quality dimensions: (1) an empirical approach through surveys amongst data consumers (e.g. MIT group), (2) an ontological one by trying to understand how information quality errors are generated (e.g. Wand and Wang), and (3) an intuitive approach, i.e. identifying dimensions on the basis of common sense and practical experience (e.g. Batini and Scannapieco, (2016) (n 40), 37).

Floridi. [51] The summary is representative of some of the most frequently mentioned information quality dimensions, but it is not an exhaustive list of all the dimensions that have been mentioned in the relevant literature on the topic. It confirms that data quality contains quite a few dimensions, beyond accuracy and completeness, e.g., 'relevance', representation, 'freedom from bias', 'level of detail', etc.[52]

The said summary should be treated as one example of the different data quality dimensions and their categorisation, because there is *no one universally accepted set of dimensions* relating to data quality or an agreement on the precise meaning and substance of each dimension.[53] As noted by Liu and Chi, there is a gap in literature, because there is no 'generally accepted DQ (data quality) model'[54] and the different approaches to defining data quality 'often create divergent and confusing definitions of basic DQ attributes.'[55] They argue that the existing approaches to derive and classify the different quality dimensions suffer from one major limitation. Namely, they are intuitive and empirical and lack a theoretical underpinning, i.e., there is no 'theoretically sound definition of DQ.'[56]

Their reference to 'theoretically sound' relates to their argument that 'data are the reflection of real-world objects through a theory that designates a set of models, methods, techniques, approaches, and heuristics used for data collection, organisation, presentation and application.'[57] Hence, 'data can only exist and have meaning through a theory'[58] and 'objective truth' can exist only within a theory.[59] Consequently, they define data quality as 'the extent to which data meet the needs and specifications of the theory.'[60]

Moreover, they note that data quality is not a static notion, but rather an 'evolutionary construct',[61] because the quality of individual pieces of data might change throughout the different stages of the data processing. This is because, as Liu and Chi have observed, data undergo changes in the course of the data evolution life-cycle (DELC)[62] or what in EU data protection terms would be called 'data processing'.[63] Liu and Chi propose that a DELC is composed of four stages: (1) data collection, i.e. capturing the data on the basis of a certain model which reflects a certain problem, and through a certain technique; (2) data

---

[51] See table in Floridi and Illari (n 36), 7, as taken from Y. W. Lee, D.M. Strong, B.K. Kahn and R. Y. Wang, 'AIMQ: A methodology for information quality assessment' [2002] 40 (2) Information & Management, 133–146.
[52] Ibid.
[53] Batini and Scannapieco (2016) (n 40), 41; Illari (n 45) 289.
[54] Liping Liu and Lauren N. Chi, 'Evolutional Data Quality: A Theory-Specific View' [2002] Proceedings of the Seventh International Conference on Information Quality (ICIQ-02), 292. 'DQ' stands for 'Data Quality'.
[55] Ibid, 294.
[56] Ibid, 294.
[57] Ibid, 292.
[58] Ibid, 295.
[59] Ibid, 301; Karst (n 34), 355.
[60] Liu and Chi (n 54), 292.
[61] Ibid.
[62] Ibid, 295.
[63] GDPR, art 4 (2); LED, art 3 (2); Regulation 2018/1725, art 3 (3).

organisation, e.g. into a model or representation; (3) data presentation, e.g. the utilised layout or format and (4) application, e.g. according to an analytical algorithm, method, model, etc.[64] These DELC stages could also be seen as the different components of what Hallinan and Borgesius refer to as the data processing framework (Section 2). Thus, the processing operations performed on a piece of data may modify it and cause different quality issues. Relevant examples are 'measurement errors during data collection, data entry errors during data organization, and interpretation biases for data presentation. In other words, the quality of captured data may not be the same as that of organized data, of presented data, and of utilized data.'[65]

Liu and Chi's theoretical proposition can be illustrated by the following example. Recently, in Denmark a problem was reported with the processing of location data as derived from mobile telecommunications data for evidence purposes in the framework of law enforcement procedures. The problem was namely that the data, as collected and submitted by the different telecommunication providers in Denmark to the police, had different formats, i.e., different data organisation and presentation formats. After receiving them, the police converted them into another, uniform format, i.e., re-organisation and re-presentation of the collected raw data. The conversion process produced mistakes in the calculation of the location data of the concerned mobile devices. In addition, it sometimes submitted the converted data to the investigators before they were fully converted. As a result, some data were lost and the data sets were incomplete, resulting in the wrong calculation of the location of suspects in some cases. This led to the necessity to re-open some of the concerned criminal cases/investigations.[66] This case exemplifies how the raw data, which seem to have been collected and stored accurately by the telecom providers, may produce incomplete and inaccurate output about the location of mobile devices. This is because the data processing technology and the logic/rules underlying it suffer from technical problems. Thus, the accuracy of the initial, raw data, did not guarantee in itself the quality of the output because of the mistakes in the processing.

The Danish example and Liu and Chi's proposal resonate with the point made by Hallinan and Borgesius that the framework for data processing as such, e.g., their analytical interpretation or *in casu* their conversion and re-calculation, should be and often is subject to certain pre-established processing standards.[67] The case also illustrates that, as the case might be, such standards could be rather technical, as is presumably the case with the Danish example. For that reason and because they are often case-specific and might differ throughout the different DELC stages, the paper argues that their details cannot always be found in data protection law: they might stem from sectoral laws or from technical standards and other disciplines, e.g., medical, biological, etc.

---

[64] Liu and Chi (n 54), 295-7.

[65] Ibid.

[66] Jesper Lund, 'Danish data retention: Back to normal after major crisis,' EDRi, 06 November 2019, <https://edri.org/danish-data-retention-back-to-normal-after-major-crisis/> accessed 16 March 2020.

[67] See Section 2 above.

The discussion so far could be summarised as follows: personal data quality should be seen as a legal and technical requirement, especially when it concerns the *processing* of personal data, starting with the collection thereof. Data accuracy, as a data protection principle, could be seen as the accuracy of the initially collected data, e.g., their correct spelling, and the accurate outcome of the processing in the sense that the output data reflect reality. The paper argues that the accuracy of the input and output data, as well as the quality of the data processing framework, should constitute in total data quality as a legal concept, where accuracy is only one element of it. The discussion, in building on Liu and Chi's model, as well as Hallinan and Borgesius's work, is visualised in Figure 1 below.

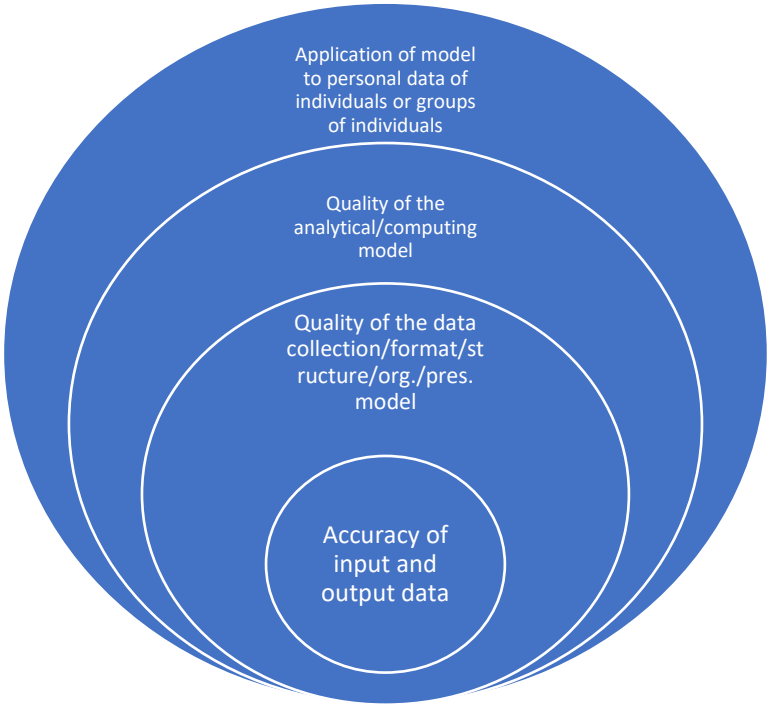

Figure 1: Data quality as a legal concept

The discussion so far also confirms Floridi's suggestion that '(t)he idea that IQ (information quality) is a *multidimensional* concept, with accuracy as only one dimension, is now embedded.'[68]

The following three sections, through examples of data protection problems, will illustrate how these can be treated as an expression of the quality issues during the different data processing stages, as represented in Figure 1 above, and how these data quality issues relate to data protection law. First, the problems of data formatting, organisation and presentation will be discussed (Section 5), then the problems of algorithmic risk profiling will be studied (Section 6) and then the less discussed topic of human cognition as a factor in ensuring quality will be analysed (Section 7).

5. Personal data quality at the stage of data collection, organisation, presentation

The discussion begins with Liu and Chi's first three stages of the DELC, from a data protection perspective.

According to the data protection principle of data accuracy in relation to personal data collection (Stage 1 DELC), it is necessary that the data be collected legally and accurately. An example to the contrary is *Khelili v Switzerland* [69], in which case the ECtHR noted that the assumption made by the Swiss police that Ms Khelili is a prostitute, on the basis of the business card found in her possession, could not be proven and therefore the assumption should be considered to be wrong. However, the risk existed that the wrong data, which were recorded as a fact in the police database, could be transferred to other police departments, which would in turn lead to the (incorrect) assumption that the information is a correct fact.[70]

In practice, problems have been reported also in relation to data presentation. A relevant example is the CJEU preliminary ruling case of *U v Stadt Karlsruhe*[71]. According to the facts of the case, the international passport of the applicant (named 'U'), as issued by the Karlsruhe administration in Germany, included not only the applicant's forenames and surname, as required by Council Regulation 2252/2004, but also his name at birth, which is not part of the personal name under German civil law. The name at birth was included in the passport in the field next to his surname and it was introduced by the abbreviation 'GEB', which means 'at birth' in German. 'GEB' was not translated in another language. This proved to be confusing for foreign administrations. Thus, when foreign authorities issued the applicant visas, they sometimes assumed that 'GEB' was actually his name and would enter 'GEB' in the name field on the visas issued by them.

This inconvenienced the applicant during checks at foreign borders, as there was a *mismatch or inconsistency* between his passport data and visa data, because the visas

[68] Floridi and Illari (n 36), 6.
[69] *Khelili v Switzerland*, App. No. 16188/07 (ECHR, 18 October 2011).
[70] Ibid. Unsurprisingly, the ECtHR ruled that the police, by not replacing the profession of Ms Khelili to 'tailor', had violated Article 8 ECHR on the human right to privacy.
[71] C-101/13 *U. v Stadt Karlsruhe* [2014] ECLI:EU:C:2014:2249.

would not accurately present the applicant's name.[72] The case demonstrates that although the personal data of the applicant were *per se* factually correct, i.e. there were no spelling mistakes and the names indeed belonged to the applicant, they were presented in a misleading format, which led to their wrong interpretation and inaccurate further processing which resulted in inconveniences for the data subject. Unsurprisingly, the CJEU ruled that '*the form in which the various components of the name of the holder appear must be free of any ambiguity and, therefore, of any risk of confusion*.'[73] In other words, administrations are required to state clearly in the captions of the name fields that the birth name is entered there[74] in order to ensure that foreign administrations will understand what data the field contains.[75] This case exemplifies the necessity for correct formatting or layout of the presentation of personal data, which is Liu and Chi's stage 3 DELC and which according to their model needs to fulfil the quality requirement. At the same time, the example clearly demonstrates that the problem of the quality of data representation is not a purely technical problem, but also a legal one, which confirms that it is reasonable to analyse data quality also as a legal construct.

Interestingly, in that case, the CJEU did not make a reference to the, then in force, Directive 95/46/EC. Rather, the CJEU and the Advocate General argued that the inaccurate presentation of the name was in breach of Article 7 Charter of Fundamental Rights of the EU (CFREU) on the fundamental right to private life, since one's own name is a constituent element of one's identity and private life.[76] Data formatting in relation to the name could be problematic 'in so far as it may give rise to doubts as to his real identity, the authenticity of the passport or the veracity of the information contained in it.'[77]

One can only speculate why the applicant in the case did not evoke the data accuracy principle and the right to rectification under Directive 95/46/EC.[78] One reason could be that the applicant did not have confidence that the data protection notion of data accuracy could be applicable in this case. If this was the reason, then it is clear that there might be a rather narrow understanding of the notion of 'data accuracy' as a data protection term. Therefore, a move towards acknowledging more explicitly the notion of 'data quality', which encompasses data accuracy, but also includes the quality elements related to data processing, might offer more flexibility when examining data processing inaccuracies and claiming their rectification.

---

[72] Ibid.

[73] Ibid, para 44. Emphasis added.

[74] Further to that end, the 'caption must be drafted in the official language of that State, if necessary accompanied by a translation, in italics, into one of the languages designated in that provision' (Ibid, para 45), which was not the case with Mr. U's passport (Ibid, para 46-7).

[75] Ibid, para 46.

[76] Ibid, para 48; also Opinion of Advocate General Jääskinen in *U v Stadt Karlsruhe* [2014] ECLI:EU:C:2014:296 para 25.

[77] C-101/13 *U. v Stadt Karlsruhe* [2014] ECLI:EU:C:2014:2249, para 50.

[78] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31, art. 6 (1) (d) and 12(b).

Finally, data organisation (Stage 2 DELC) is also essential in ensuring the correct attribution of information on the same person in a database or across databases. In that respect, mistakes have been reported in relation to the Schengen Information System, such as entering a name instead of a document number or date of birth entered in the name field.[79] In another example, biometric data of one person were attached to the alphanumeric data of another one.[80] This means that *incorrect links* were made between different sets of correct data. The significance of this issue is expected to grow in the following years when the interoperability between the databases in the Area of Freedom, Security and Justice (AFSJ) is implemented. This is because one of the elements of interoperability would be the Multiple Identity Detector (MID), pursuant to which links between similar data, as stored in the different AFSJ databases, will be created. For example, red links will be created for similar but different data which are believed, e.g., to belong fraudulently to the same person.[81] Should mistakes occur when linking the data, this would arguably be a breach of the data quality principle because as Wang and Strong argue, the correct identification of the same person across different databases is perceived as part of data and system quality.[82] This sounds reasonable because even where the underlying data are correct, but the link between them is wrong or the correct personal data are attributed to the wrong person, then there is a systemic problem. That is why the accuracy of the data is not enough to guarantee accurate outcomes, i.e., correct person identification.

---

[79] European Commission, 'Commission Staff Working Document Accompanying the document: Report from the Commission to the European Parliament and the Council on the evaluation of the second generation Schengen Information System (SIS II) in accordance with articles 24 (5), 43 (3) and 50 (5) of Regulation (EC) No 1987/2006 and articles 59 (3) and 66 (5) of Decision 2007/533/JHA {COM (2016) 880 final}' SWD (2016) 450 final, Brussels, 21 December 2016.

[80] European Union Agency for Fundamental Rights, 'Fundamental Rights and the interoperability of EU information systems: borders and security' May 2017, 8 and 34; European Commission, 'Commission Staff Working Document: Evaluation of the implementation of Regulation (EC) No 767/2008 of the European Parliament and Council concerning the Visa Information System (VIS) and the exchange of data between Member States on short-stay visas (VIS Regulation) / REFIT Evaluation Accompanying the document Report from the Commission to the European Parliament and the Council on the implementation of Regulation (EC) No 767/2008 of the European Parliament and of the Council establishing the Visa Information System (VIS), the use of fingerprints at external borders and the use of biometrics in the visa application procedure/REFIT Evaluation {COM(2016) 655 final} {SWD(2016) 327 final}' SWD (2016) 328 final, Brussels, 14 October 2016, 54; European Union Agency for Fundamental Rights, 'Under watchful eyes: biometrics, EU IT systems and fundamental rights,' (2018), 15-16.

[81] Regulation (EU) 2019/818 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration and amending Regulations (EU) 2018/1726, (EU) 2018/1862 and (EU) 2019/816, (2019) OJ L135/85; Regulation (EU) 2019/817 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of borders and visa and amending Regulations (EC) No 767/2008, (EU) 2016/399, (EU) 2017/2226, (EU) 2018/1240, (EU) 2018/1726 and (EU) 2018/1861 of the European Parliament and of the Council and Council Decisions 2004/512/EC and 2008/633/JHA, (2019) OJ L135/27 (Interoperability Regulations), Chapter V.

[82] Wang and Strong (n 44).

6. Data quality at the stage of data interpretation, profiling, and decision-making

In Stage 4 of their DELC model, Liu and Chi refer to the application of a certain data processing model, which could also be understood to mean an algorithm. They do not refer explicitly to profiling or automated decision-making, which is one of the widely discussed topics in data protection scholarship and regulatory debates. The Article 29 Working Party identifies that in the profiling and automated decision making process, next to the data collection and initial analysis, there is the stage of the abstract profile construction and its subsequent application to the personal data of an individual.[83] Following Liu and Chi's work, as well as Hallinan and Borgesius's argument, the profile construction and application should be subject to quality requirements, although Hallinan and Borgesius refer to accuracy here.

For example, the 2019 Finnish Presidency noted in relation to the quality of algorithmic decision-making that 'the transparency and *correctness of algorithms* used in all applications of artificial intelligence as well as other appropriate safeguards need to be looked at in order to maintain the ability to *verify the credibility of the results proposed and to ensure the overall accountability and lawfulness of such algorithms*.'[84] Although the note refers to 'correctness' and not to quality, what is more important is that it pays attention to correctness in relation to *algorithms*, not simply to the input data. This confirms the importance of the data processing framework for the legal operation of algorithms, whether the different stakeholders prefer to term it as 'correct' or 'of good or high quality.' Similarly, the Article 29 Working Party calls for quality assurance checks of the decision-making systems in order to ensure that these are fair and do not result in discriminatory results.[85] This is necessary, because as Malgieri and Comandé argue, in the framework of algorithmic decision-making mistakes could happen not only with regard to the input data, but also with regard to the design of the algorithm. These could be due to machine biases embedded in the methodology and structure of the algorithm.[86] The major risk in such cases is the possible incorrect assessment of the data, the associated data analysis, and the consequently incorrect results.[87] To avoid such risks, one needs to understand not only the data, but also the biases in the technology and the limits of the data and the technology.[88]

In data protection law there are requirements relating to *non-bias in profiling and automated decision-making cases and requirement on non-discrimination*, which link to the requirements identified in the interdisciplinary discussion, namely data being free from

---

[83] Article 29 Working Party Guidelines (n 25), 12.

[84] Council of the European Union, 'Note from the Presidency, The future direction of EU internal security: new technologies and internal security - Preparation of the Council debate' 12224/19, Brussels, 18 September 2019, 7, emphasis added.

[85] Article 29 Working Party Guidelines (n 25), 32.

[86] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' [2014] (7) (4) International Data Privacy Law. They consider the other articles, too (GDPR, Art. 13-14, 22 and recital 71).

[87] Hoeren (n 15), 35 and 37.

[88] Hoeren (n 15), 35. Also Bygrave (2002) (n 42), 308.

bias, and also to the requirements from the reviewed computer science literature on the quality of the models and the data processing operations at large. The EU data protection framework contains few requirements concerning decision-making, especially profiling and automated decision-making, from which the requirements on non-discrimination and non-bias can be more or less directly derived. For example, the LED prohibits discriminatory profiling based only on special categories of data, and requires that automated decisions based on such data be accompanied by adequate safeguards and that automated decisions and profiles should have appropriate legal basis.[89] The GDPR also requires certain safeguards in relation to automated decision-making and profiling be put in place, including when the processing is based on special categories of data.[90]

Furthermore, the GDPR contains a clarification in Recital 71, namely that the controller should ensure that when automated decisions are made, 'factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised…' *inter alia* through safeguards such as 'appropriate mathematical or statistical procedures for the profiling.'[91] The wording 'factors' is relatively broad and could be argued to mean the abstract criteria used for the profiling, the selection of the input data, etc. In addition, Recital 71 GDPR recommends that a result is achieved, namely that the 'risk of errors' is reduced to the minimum. This is to be achieved, amongst others, through scientifically proven methods and formulas.

Although the GDPR and the LED do not contain a full catalogue of requirements related to profiling and automated decision-making, the above examples demonstrate that requirements on the *quality of the processing*, not simply on the accuracy of the (input) data, do exist in EU data protection law, although the term 'quality' is not mentioned. Furthermore, their non-exhaustive nature and similarity to the technical requirements on data processing quality in other disciplines might signal that good quality of the processing is essential for legal compliance when personal data are processed, but that the exact technical requirements should be developed by or adopted from the more technical disciplines. This suggestion seems to be strengthened by the draft AI Act,[92] which seems not to be able to exhaustively regulate the quality of processing requirements in relation to one particular set of technologies, namely AI ones, but which nevertheless clearly demonstrates that data quality is more than the accuracy of input data (see Section 3 above).

For example, when Passenger Name Record (PNR) data are processed in order to identify passengers who could be engaged in serious crime, not only must the input PNR data be accurate, but also the logic of the profiling algorithm and the result of the data processing operation. This means that the algorithm should target only those passengers who would indeed be reasonably suspicious, as the CJEU pointed out in the *Canada PNR* Opinion.

---

[89] LED, Art 11 (1) – (3).

[90] GDPR, Art 22 (1), (3) – (4).

[91] GDPR, recital 71 and Winfried Veil, 'Artikel 16 Recht auf Berichtigung' in Sibylle Gierschmann, Katharina Schlender, Rainer Stentzel and Winfried Veil (eds), *Kommentar Datenschutz-Grundverordnung* (Bundesanzeiger Verlag 2018), 495, para 4.

[92] Proposed EU AI Act.

Specifically, the Court concluded that *the extent of the interference* of the *automated processing* of PNR data with Articles 7 and 8 CFREU 'essentially depends' on the pre-established criteria and models, as well as on the background databases against which the processing is based.[93] Therefore, quality requirements should be established in relation to the profiling criteria and models. These requirements should ensure that the criteria and models are 'specific and reliable' in order to identify targets against whom there is a 'reasonable suspicion' of being involved in serious crime.[94] From the judgment, the following three major conclusions could be derived. First, quality is important for complying not only with data protection law, but also for complying with the CFREU, similarly to the conclusion in *U v Karlsruhe*.

Second, it is notable that it is *the Court* that established that data processing, *in casu* risk profiling, should be of adequate quality and then set out high-level quality requirements. This again confirms that data quality should not be seen as a purely technical concept, even though the detailed quality requirements are more likely to be developed and implemented on a technical level.

Third, the CJEU also ruled that the abstract model should be designed in such a way as to fulfil the purpose of the processing, i.e., to solve the real-life problem, *in casu* detection of serious criminals. This clearly links to the purpose limitation principle in data protection law, pursuant to which before personal data are processed, starting with their collection, the purposes should be clearly defined.[95] Read together with the data minimisation requirement (data should be 'adequate, relevant, and limited, to what is necessary in relation to the purposes for which they are processed')[96] and the data accuracy

---

[93] Case *Opinion 1/15 of the Court* (Grand Chamber) [2017], ECLI:EU:C:2017:592 (Canada PNR Opinion), para 172; Niovi Vavoula has noted that sometimes the background databases used for risk profiling are flawed, i.e. the data storage and further processing might be in breach of EU data protection law, see Niovi Vavoula, 'Artificial Intelligence (AI) at Schengen Borders: Automated Processing, Algorithmic Profiling and Facial Recognition in the Era of Techno-Solutionism' [2021] *European Journal of Migration and Law* (forthcoming).

[94] *Canada PNR Opinion*, para 172.

[95] GDPR, Art 5 (1) (b); LED, Art 4 (1) (b); Regulation 2018/1725, Art 4 (1) (b).

[96] GDPR, art 5 (1) (c); LED, Art 4 (1) (c); Regulation 2018/1725, Art 4 (1) (c); See in particular two ECtHR cases on data minimisation. For example, in the *P.T v the Republic of Moldova* case, the Moldovan authorities had entered on the military service certificate of the applicant that he had not done his military service because he was HIV-positive. This certificate needed to be presented often in front of different administrations, e.g. when applying for an ID card, which is obligatory in Moldova. Non-surprisingly, the ECtHR ruled that entering such excessive information as the HIV-positive status on the military service certificate constituted an interference with Article 8 ECHR. See *P.T. v the Republic of Moldova*, App No. 1122/12, (ECHR, 26 May 2020). As the Court noted, 'the need to include such a degree of sensitive medical details in a certificate which could be requested in a variety of situations where the applicant's medical conditions was of no apparent relevance, such as when applying for employment …' was not motivated (para 32). In another case, *Stavropoulos and Others v. Greece*, the contested issue was that in Greece there was a widespread practice that registration authorities would put a note 'naming' on the birth certificates next to the name of new-born children who have not been christened. This addition was not prescribed by law and in effect revealed religious information about the concerned persons. See *Stavropoulos and Others v. Greece*, App. No. 52484/18, (ECHR, 25 June 2020). Note that the case was examined under Article 9

requirement (data should be kept accurate an up to data in relation to the purposes of the processing),[97] this means that the risk assessment model should be designed in such a way as to collect personal data which are relevant, adequate, accurate, current, and necessary, for this particular purpose, so that the technology achieves its pre-defined purpose. This echoes Liu and Chi's proposition that data quality and data meaning are determined by a 'theory', e.g., by the methods and models of processing, which are a reflection of a real-life problem. The impact of the change of this 'theory' or context was exemplified in *Google Spain*,[98] where information that the applicant's house had been put up for auction so that they could pay off their debts had been published in the local newspaper. Two decades later, the news article was still accessible when the name of the applicant was searched in Google and the information therein continued to influence the perceptions people had about the applicant, even though the debts had been paid off. This exemplifies how the further usage of the data, beyond the fulfilment of the original purpose of the processing, *in casu* the auction had taken place, could create impressions of individuals which could influence decisions about them, e.g., whether to employ them or not, and which could be inadequate and irrelevant if they are outdated and wrong.[99]

Going back to the example of risk profiling, to achieve good quality, e.g., the accurate identification of suspects through data analysis based on pre-established models, one first needs to understand the problem which the data processing aims to solve. Put succinctly by Bygrave, 'the quality of any information (and any data or IS (Information System)) can only ever be fully assessed in the light of the models upon which the information is based. Moreover, poor conceptualisation of a problem (or what might be termed poor model quality) will tend to result in poor interpretation and application of the information which is processed to address the problem.'[100] This observation relates to Liu and Chi's proposition that a real-life problem could be modelled in different ways and that the quality of the chosen model should be measured in relation to how well it represents the given problem.[101] *A fortiori*, the conceptualisation of the solution to the problem, e.g. the profiling algorithm for identifying passengers who might pose security risks, should be preceded by a clear and precise definition of the problem and fulfil the applicable quality requirements. Otherwise, the mere collection of information will not achieve its purpose if the model for the interpretation of the data in relation to the identified problem is not of good quality.[102]

ECHR on the right to freedom of religion. Having found a violation of that provision, the ECtHR did not deem it necessary to examine it also under Article 8 ECHR. However, one could doubt that the Court would have considered the addition of the contested information as legitimate and/or necessary and proportionate (in addition to the fact that it was not prescribed by law).

[97] GDPR, Art 5 (1) (d); LED, Art 4 (1) (d); Regulation 2018/1725, Art 4 (1) (d).

[98] C-131/12 *Google Spain* [2014], ECLI:EU:C:2014:317.

[99] Ibid, para 94.

[100] Bygrave (1996) (n 35) 'IS' stands for 'Information System.'

[101] Liu and Chi (n 54) 294.

[102] Batini and Scannapieco also conclude that the efficiency or effectiveness of processes or decisions which are based on information can be assessed only in relation to the problem which the information and its processing seek to solve (See Batini and Scannapieco (2016) (n 40), 351).

In that respect, computer scientists have noted that whereas a certain system which is applied to a real-life problem could be initially accurate, this accuracy refers to the static training data.[103] At the same time, the real-life problem, which the static model is applied to, could change, e.g., the behaviours of the people that the system seeks to evaluate could change, and thus the static model is likely to inaccurately assess the population in the future, i.e., after the change. In computer science, this concept is referred to as 'concept drift.'[104] Hence, it can be concluded that 'accuracy is not a static measure'.[105]

Finally, even if the chosen (algorithmic) model accurately reflects the problem, leads to the sought solution and evolves together with the problem it seeks to solve, one significant issue remains. Algorithms which aim to predict the behaviour of individuals in the future cannot be designed with 100% accuracy because of their predictive nature and the lack of a truth as a baseline for comparison.[106]

All the quality problems as discussed so far should not be seen as purely technological: humans, their understanding of the world ,and their interaction with technologies, can also influence the quality of the processing.

7. Human cognition: An underdiscussed factor in data quality

The discussion so far has been focused on a variety of technical and legal issues related to data quality, which this paper argues to include data accuracy and quality of data processing requirements. However, relatively little attention is paid to the role of human cognition in ensuring good quality processing and results. Bygrave has argued that 'the set of factors relating to human cognition play a relatively large role in determining data (…) quality.'[107] Examples from the 1990s include poor thinking, interpretation, and application, e.g., designing wrong models to understand the problem at hand.[108] The human factor calls not only for a high cognitive ability of individuals to understand information.

Nowadays, with the growing usage of algorithms, cognitive factors could potentially also refer to the ability of human decision-makers to oversee the technology, to understand how these algorithms work, to assess the produced result, and to independently confirm

[103] Indre Zliobaite, 'Learning under Concept Drift: an Overview' [2009] Technical report, Vilnius University, https://arxiv.org/abs/1010.4784 accessed 03 April 2020; Reuben Binns and Valeria Gallo, 'Accuracy of AI system outputs and performance measures' [2019] ICO Blog, <https://ico.org.uk/about-the-ico/news-and-events/ai-blog-accuracy-of-ai-system-outputs-and-performance-measures/> accessed 03 April 2020. Please note that Zliobaite refers to 'accuracy' and 'accurate', and not to 'quality'. This could signal the fact that Zliobaite might accept that these concepts are the same and bearing in mind the substantive similarity between these concepts, this is not a problem. It simply means that Zliobaite's reference to 'accuracy' does not automatically mean that model accuracy could not mean model quality.

[104] Ibid.

[105] Binns and Gallo (n 103).

[106] Article 29 Working Party Guidelines (n 25), 17 – 18.

[107] Bygrave (2002) (n 42), 106.

[108] Bygrave (1996) (n 35).

or challenge the end result.[109] Human cognition could also refer to the ability of the person assessing the input data to interpret them correctly, without distorting the facts. One of the reasons for distortion of facts, otherwise referred to as a faulty interpretation, could be the wrong understanding of human beings in their capacity as decision-makers.[110]

An important aspect of human cognition is also the ability of those working with the data and data processing technologies to understand that data-driven technologies and solutions might not always be perfect. In other words, data users should be aware of the (potential for) limitations of the data they are using[111] and should therefore accept that data processing technologies do not produce and represent an ultimate truth, but are prone to failures similarly to human beings. This awareness is important in order to allow those who work with the technologies to think critically of the functioning and results of these technologies and to understand that data might not represent accurately the reality in relation to the different problems in the world. Otherwise, there is a risk that those who work with technologies will lose the ability to think of the world outside the framework of the data and data processing technologies and the results they produce. This could in turn make it more difficult for the concerned individuals, i.e., data subjects, to convince them that the output of the technology might be incorrect.

Furthermore, human cognition is also related to the *harmonised* understanding of the available information in the framework of databases which are used by a variety of authorities and officers from different countries, as is the case with most EU-wide AFSJ information systems. A problem might be that the different users might not use the same terminology or data classification. As Bygrave has exemplified with a real case from the 1990s, a Swedish municipality generated many wrong hits when trying to identify people in illegal receipt of housing aid, because one of their criteria was income, which had a different meaning across the different databases used for the profiling.[112] Further issues may emerge, e.g. in the framework of the EU PNR framework, pursuant to which Member State law enforcement authorities may exchange raw data and reach risk assessment conclusions on risky individuals.[113] Problems could emerge when the person making the evaluation and the recipient of the evaluation may not share the same language or understanding of the facts and the elicitation of evaluations from facts.[114]

Then, even if the different users develop a similar terminology, this might lead in turn to another problem, namely it might create a language barrier between the users and non-users,[115] e.g., between data controllers such as governments, and data subjects, e.g., citizens. One could argue that this might come at the expense of transparency towards the

---

[109] That is why the Proposed EU AI Act anchors requirements of transparency towards the users of AI technologies, esp. Art. 13 and 14.

[110] C-266/05 P *Jose Maria Sison v Council of the European Union*, [2007] ECLI:EU:C:2007:75, para 67.

[111] Kieron O'Hara, 'Enhancing the Quality of Open Data,' in Floridi and Illari 2014, (n 36), 208.

[112] Bygrave (2002) (n 42), 106.

[113] Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime [2016] OJ L119/132 (EU PNR Directive), Art 9 (1).

[114] Karst (n 34), 356.

[115] Ibid, 361.

data subjects and supervisory authorities. This illustrates the necessity for explainability of the results of the processing to the concerned individual in an intelligible form, 'using clear and plain language.'[116] In other words, the efforts to achieve good quality should not result in such a complexity of the processing, which will make it unaccountable.

To sum up, the above overview revealed the main themes that have emerged in literature with regard to data and information quality. These are namely the accuracy of the individual pieces of data, the quality of the data processing operations and the model used for the processing, such as criteria for profiling, and the final result, e.g., correlations discovered between data, as well as human understanding. The discussed quality requirements can be found not only in technical literature, but they are also starting to be set out in legal sources. For example, data quality requirements can be derived from EU data protection law and case law. It would be then logical that individuals have means to enforce the quality requirements when their data are processed. The next section will examine whether and in how far the right to rectification could be evoked by individuals in order to correct data quality issues.

8. The right to rectification should be interpreted broadly

If the paper's proposition - that data quality can be seen as a legal concept which includes but is broader than the accuracy of input and output data - would be accepted by data protection professionals, then logically the question arises whether this 'new' principle influences the scope of the right to rectification.

The provisions on the right to rectification in EU secondary law do not seem to take account of *data processing quality*, because, strictly speaking, they refer to the rectification of inaccurate personal data and the completion of incomplete data.[117] The scarcity of case law on the right to rectification also does not help to answer the question whether individuals can evoke the right to rectification to address the broad range of issues related to the quality of the processing of the data. The following paragraphs will explore the impact which the broad interpretation of the data quality principle could have on the scope of the right to rectification and will also point out the questions and uncertainties which the principle of data quality poses for this right.

The first step to being able to exercise of the right to rectification is the ability of the data subject and the data controller to discover the mistake in the data and/or in their processing, which has produced inaccurate outcomes. Ironically, wrong data may themselves impede the discovery of the mistakes in them. This could be the case where, e.g., the names under which information on a person are stored in a database are misspelt, e.g., because of transliteration mistakes. This is because if a person requests access to their personal data in order to verify the accuracy of data stored about them and possibly request corrective actions, the search for these data will likely be performed with the

[116] GDPR and LED, Art 12 (1).
[117] GDPR, art 16 and LED, art 16 (1).

accurate data, e.g., with the data directly presented by the data subject. However, a possible misspelling of the name(s) in the database, as stored by the data controller, might lead to non-location of the (wrong) data of the individual. This could also result from poor data management. Thus, even if a file exists on them with the requested controller, the data subject and the data controller might remain unaware of this; hence the importance of sound data management practices and complying with data quality standards, not only in order to obtain accurate outcomes of the processing, but also to fulfil their obligations in relation to enabling the exercise of data subject rights.

The second step is the actual rectification of the source and result of the identified problem. To what degree the right to rectification could solve the identified problem, however, depends on whether a broad view on its scope is accepted. The paragraphs below will focus on three cases, which represent quality issues related to the different stages of *data processing*, as identified in the previous sections, in order to demonstrate how the right to rectification could be read broadly to rectify such data processing issues. These three cases are: (1) inferential data and the underlying logic leading to them; (2) biometric data and (3) formatting, structure and organisation.

a. Inferential data and opinions, and the processing from which they resulted (Case 1)

This category of data is relatively broad and shares one common trait – it is non-factual data. It could encompass risk assessments, such as those in the framework of PNR, witness statements in criminal proceedings, credit worthiness, etc. In the data protection instruments, the only reference which explicitly restricts the right to rectification concerns the justice sector. The LED clarifies that 'the right to rectification should not affect (…) the content of a witness testimony.'[118] This situation could be explained by the fact that participants in judicial proceedings cannot and should not be allowed to force witnesses to give a particular statement or to change their statement, as witness statements should be independent. However, the LED clarifies that a witness statement should be clearly marked as such.[119] This goes to the core of the principle of separating opinions from facts in the LED.[120]

Similarly, in the non-law enforcement context, the ICO has argued in its guidance on the GDPR that the adequacy of opinions should refer not necessarily to the correctness of the content of the opinions, but rather to the correct indication that certain information is an opinion and not a fact. Furthermore, information should be added to help interpret the opinion correctly.[121] Such a clear marking, where it might be missing, seems to be a proper measure to rectify potential misunderstandings about the significance and value of

[118] LED, recital 47.
[119] LED, recital 30.
[120] LED, art 7 (2); Regulation 2018/1725, art 74 (2). See also Principle 3.2. in Council of Europe Recommendation R (87) 15.
[121] Information Commissioner's Office, 'Principle (c): Data minimisation' https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/ accessed 11 March 2020. Note that this remark was made in the framework of the data minimisation principle, not the data accuracy principle.

information. Thus, distinguishing facts from opinions under the GDPR and the LED should be treated as a rectification measure. In the literature above it was also suggested that the degree of accuracy of the data should be indicated. This requirement is explicitly mentioned in the draft AI Act[122] and in the law enforcement sector, which means that it would apply, for example, in the framework of PNR profiling.[123] It is not explicitly mentioned in the GDPR, though. A crucial point concerns the enforcement of this requirement. The problem is that no standards on the categories or degrees of accuracy of inferential data seem to exist. Even if there were any standards, it is not immediately clear whether a data subject could, pursuant to the right to rectification, request that the data controller reassess the degree of accuracy and mark this degree differently, i.e., rectify the metadata concerning their personal data.

The discussion so far illustrates the potential role of the right to rectification as a measure for accuracy, which could help the quality management of data and increase the transparency about its reliability towards the users, i.e., the controllers and processors. However, the question arises as to whether clearly distinguishing between inferential data and factual data, and annotating their degree of accuracy, is the only corrective measure in relation to opinions and inferences. The question is very pertinent, bearing in mind the fact that this paper has demonstrated that quality requirements exist in relation to the framework for processing of the data, including in profiling cases.

The below paragraphs will discuss how one should interpret and apply the right to rectification in relation to (1) the abstract criteria, (2) their application to an individual and (3) the final result, i.e., the inference. This focus was chosen, because although it was argued in the previous sections that legally speaking quality applies to the processing of data, e.g., to the profiling algorithm, it has not been discussed how the right to rectification could apply to them, beyond the fact that the raw personal data, as collected, have to be accurate.

With regard to the abstract criteria (1), the first observation is that they in principle do not relate to a specific individual until they are applied to the personal data of individuals. By contrast, pursuant to the right to rectification in data protection law, the data subject may claim their right to rectification in relation to inaccurate data 'concerning him or her', and as the case might be – to their completion.[124] This implies that the data subject may not request the rectification of the abstract criteria, according to a strict reading of the wording of the right to rectification. Nevertheless, as discussed above, one element of data quality identified in the interdisciplinary overview is that data should be necessary, relevant and adequate for the purpose of the processing.[125] However, if the abstract profiling criteria are designed in such a way as to collect personal data which are not relevant, representative and adequate for the said profiling purposes, but instead omit the collection

[122] Proposed EU AI Act, art. 13 (3) (b) (ii).
[123] See EU PNR Directive, recital 27j; LED, art 59.

[124] GDPR, art. 16; LED, art 16(1).
[125] European Union Agency for Fundamental Rights and Council of Europe, 'Handbook on European data protection law' [2018], 125.

and further processing of relevant data, this could be interpreted as a breach of the data minimisation principle,[126] which could moreover result in inaccurate inferences. Furthermore, if the algorithm is programmed to select unrepresentative and irrelevant data, this might lead to bias; if it has been programmed to collect mostly sensitive data, which form the basis of conclusions and decisions, then it might be in breach of the rules on profiling and automated decision-making under EU data protection law.[127] The described scenario could be also argued to be in breach of the purpose limitation principle if the criteria were not selected in such a way as to fulfil the purpose of the processing. This could be the case when they target persons who have never been and never will be involved in criminal activities, and falsely label them as posing high (security) risk, such as in the PNR framework, and not individuals whose behaviour is indeed indicative of criminal intent or past criminal activity. These examples illustrate how profiling models which do not correspond to the quality requirements in the technical literature on algorithmic quality could at the same time be in breach of EU data protection law.

In the described scenario, for affected individuals to be able to have their risk status modified to low risk or no risk at all (i.e., the inference (3)), where they submit a rectification request, it might sometimes be necessary to modify the whole algorithm and not only the inference, where the controller becomes aware of the weaknesses in the model. This could be the case, e.g., where the algorithm has not been programmed to collect relevant and adequate personal data in relation to the problem the algorithm seeks to solve. As a result, the collected and analysed personal data of all or the majority of the concerned data subjects, including in relation to the person who exercises their right to rectification, are likely to be *incomplete* and produce inaccurate results because they are incomplete, which falls clearly within the scope of the right to rectification. Examining such rectification requests could lead the controller to modify the whole model, which, however, should be seen as a result not merely of a data subject rectification request, but also as stemming from the accountability obligations of the controller.[128] Whether this proposition would be followed by the courts, remains to be seen if it is tested in court.

If the proposition would not be followed by courts and the abstract criteria may not be rectified following a rectification request, there is still the possibility that a data subject could request the rectification of the *application* of the abstract criteria to their personal

---

[126] Information Commissioner's Office, 'Principle (c): Data minimisation' https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/ accessed 11 March 2020; Gary T. Marx and Nancy Reichmann, 'Routinizing the Discovery of Secrets: Computers as Informants' [1985] (1) Software LJ, 112. An example they give is the potentially incomplete knowledge on criminals since profiles are based on the basis of the knowledge collected by those who get caught and not on those who do not get apprehended; In terms of risks for the data subject, Karst has noted that the more complete the files are, the more likely it is for those who work with the data therein to assume that these contain the full profile or truth about a person. Thus, there is a danger that those who work with digital files might be misled that the collected data, and no other data, contain the relevant factors about a person and ignore or mistrust other information. See Karst (n 34), 355 and 361-362.

[127] GDPR, art 22 (4); LED, art. 11 (3); PNR Directive, art. 6(4).

[128] GDPR, art 5(2) and LED, art 4(4).

data (2). This means requesting the rectification of how their personal data have been interpreted by the algorithm and those who oversee it and take final decisions, how each element of the input data was weighed and influenced the final outcome, and the correction of the potential misinterpretation of the facts, as the case might be. Such an interpretation of the right to rectification is defensible, because the application of the algorithmic reasoning applies to an individual, it concerns their personal data, and the wrong interpretation of the data could result in inaccurate inferences concerning a specific individual. In case the controller reassesses the available data and concludes that they might not be conclusive about the fact that someone might pose a risk (in risk profiling cases), especially where the data were misinterpreted and/or wrong correlations were made in individual cases, the controller may decide to disapply the final result. In practice, this means that if the profile concluded that a person poses a high security risk, by way of rectification, the controller would correct it to 'low security risk'. Effectively, this is a correction of the inference itself ((3) above).[129] Nevertheless, before correcting the final result, ideally the controller should still examine how and why the contested result was obtained in order to check whether the problem lies in the design of the algorithm, especially in relation to its application to the data of the concerned person. This could lead to the discovery of breaches of different data protection provisions, as discussed in relation to the abstract algorithm in (1) above and its potential rectification. For such checks to take place, an important factor is the awareness of the controller that data and data processing technologies have limitations (see Section 7).

Irrespective of the outcome of the rectification request, an interesting question is how an individual may state their case that the data processing and its result could be in breach of data protection law and trigger the application of their right to rectification. In profiling and similar cases, the concerned data subject might resort to submitting a supplementary statement in which they present alternative facts to be taken into account and/or offer an alternative explanation and interpretation of the facts.[130] This is one of the novel provisions on the right to rectification as compared to Directive 95/46/EC. If the data subject indeed takes the opportunity to submit a supplementary statement and to use it as a tool for arguing with, including rebutting, the controller and the technology, then it could be seen as a sort of an adversarial relationship between controllers and data subjects.[131] The GDPR and the LED, though, do not clarify what the obligations of the controller are upon the receipt of such a supplementary statement which contains further data, contextual information, and suggestions for interpreting the situation of the concerned individual.

It could be argued that pursuant to the different data protection principles and the accountability obligations of the controller, they should examine the information in the supplementary statement and where the information is a clear indication that the result of the processing and the underlying logic, e.g., algorithm, might suffer from mistakes, then

[129] Article 29 Working Party Guidelines (n 24), 17.

[130] Ibid. For this to actually happen in practice, the data subject needs to receive information about the decision-making logic, including on the basis of which criteria a decision or profile was made and how the criteria were interpreted.

[131] Diana Dimitrova, 'Data Subject Rights: The Rights of Access and Rectification in the Area of Freedom, Security and Justice', PhD Dissertation, Vrije Universiteit Brussel, 2021.

the controller should be obliged to correct, update and complete the personal data and examine the quality of the profile. However, as of now, there is no case law on the right to rectification which confirms this interpretation.

What is becoming evident, though, is that if the right to rectification in EU data protection law is to be an effective tool for protection in the face of the new sophisticated technologies, which seek to profile and predict behaviour on the basis of abstract criteria, then it should not be seen only as a tool for upholding input data accurate. Instead, it should be read broadly, i.e., as having the potential to rectify algorithm model issues. This would be possible if the right to rectification is read together with the accountability obligations of the controller, and if it is accepted that the right to rectification can be evoked to correct not only the inaccuracy of factual, input data, but also the quality of the data processing model. Such an interpretation does not seem to be precluded by data protection law. On the contrary, the above paragraphs demonstrated that quality requirements of abstract algorithms could be derived also from the principles of data protection law.

b. Biometric data (Case 2)

Biometric data present a particular challenge, not only with regard to their collection, but also with regard to their indexing and matching. This is because their accuracy cannot be examined only in the abstract. Next to the necessity of having a good quality enrolment, i.e., collection, quality can be examined in relation to how accurately it can verify or identify data subjects' identity. This means that quality is not only about enrolment accuracy, but also about the success and reliability of the matching result when the enrolled biometric identifiers are compared to the live identifiers or the enrolled biometrics of the same person in another system.[132] This refers to the ability of the matching technology to produce correctly a biometric 'match' or a 'mismatch.' A problem could emerge, e.g., when both the enrolled biometrics and the live ones used for matching belong to the same person and are of good enough quality, but the technology wrongly rejects the match between them.[133]

The problem with exercising the right to rectification in relation to biometric data processing is that it is not clear how a person may state their case in the case of an alleged false rejection or false acceptance. For example, may the data subject request an independent assessment of the accuracy of the mismatch? This could be reasonable, since proving a biometric match accurate on inaccurate requires a complex technical comparison with the help of technology, which average citizens do not dispose of. Then, even if the mistake is proven, does the right to rectification require the controller to provide a new biometric enrolment to the data subject and/or improve their data matching technology

[132] European Commission, 'Report from the Commission to the European Parliament and the Council: The availability and readiness of technology to identify a person on the basis of fingerprints held in the second generation Schengen Information System (SIS II)' COM (2016) 93 final, Brussels, 29 February 2016, 4.

[133] Council of Europe, 'Progress report on the application of the principles of Convention 108 to the collection and processing of biometric data (2005)' T-PD, February 2005, paras 67-69.

where it is found that the enrolled data are correct, but the mismatch is due to mistakes in the matching algorithm? The case of mismatching due to the algorithm is similar to the above-discussed problem of the quality of profiling algorithms. It reinforces the argument that for the model, or more broadly – processing model, quality is an important factor in determining the accurate outcome of the processing, and that for the right to rectification to be effective, it should also be interpreted to apply to the processing model.

c. *Technical issues (Case 3)*

In *U v Stadt Karlsruhe,* the CJEU ruled that the discussed formatting problem should be rectified, e.g., by translating 'GEB' or otherwise presenting the name at birth in a more understandable way. One can imagine that in that case, the corrective measure might have a wider impact than rectifying the presentation of the data in the passport of the applicant, i.e., presumably all German passports now must have a translation of 'GEB' where a German citizen decides to include the name by birth in their international passport. This demonstrates the wider impact of the right to rectification[134] and how it could result in modifying systemic issues, as sometimes the rectification of the data in relation to one individual could be implemented only if the same issue is rectified in relation to other individuals in the same situation.

In practice, however, there are more complex technical issues than the name presentation in passports. A prime example is the EU-wide Schengen Information System, into which all Schengen Member States and most EU Member States may enter data and search the data entered by other Member States. They may also create a national copy of the System, which has to be at all times synchronised with the data on the central system. However, it has been reported that sometimes the two systems are not synchronised.[135] Presumably, this could lead to wrong (mis)matches. For instance, data of persons whose data are indeed stored on the SIS might not be located. In such a case, it is again not straightforward whether a data subject has the right to claim, on the basis of their right to rectification, that the national settings in certain Member States be changed in order to produce accurate matches, if it is presumed that they can find out about this problem. Looking at the data subject-centric phrasing of the right to rectification, such a claim seems less likely to succeed. This is problematic, because it could cause data quality issues which could have a negative impact on individuals.

Finally on organisational matters, if a law enforcement authority has not created a system for distinguishing the different categories of individuals, as required by the principle of data quality in the LED, it is again not clear what the result of the claim of a data subject to be clearly indicated, e.g., as a victim, should be. The most straightforward answer is that the status of the concerned person should be added to their personal file, e.g., that they are a victim. It is not excluded that such requests for rectification could motivate the controller to re-organise the said system and order the other available personal data into categories.

---

[134] Although, as noted in Section 5 above, *in casu* the problem was examined on the basis of the fundamental right to private life, not under the right to rectification.

[135] European Court of Auditors, 'Special report No20: EU Information Systems supporting border control – a strong tool but more focus needed on timely and complete data' 2019, 16, para 21.

However, due to the individual-centric nature of the right to rectification, i.e., individuals may evoke it only when the processing of their personal data is concerned and not in principle to correct general data organisational problems, a data subject request for reorganisation of the whole system looks less plausible. However, if a controller is aware that the data are not organised in the categories prescribed by law, then again pursuant to the accountability obligations of the controller, the controller might have to re-organise their systems; hence the indirect effect of the right to rectification on systemic issues.

9. Conclusion

The paper set out to study the relationship between the data protection concept of personal data accuracy and the notion of data quality. It demonstrated that data quality seems to have emerged as a concept outside the legal field, especially in the business and computer science fields. The paper showed that certain elements and features of data quality identified in these disciplines can be argued to have legal foundations. These can be found, e.g., in the different data protection principles related to the processing of personal data, the relevant CJEU case law, and prospective legal instruments such as the AI Act Proposal. The paper proposed that the instruments should be read to provide the legal basis not only for the principle of data input and output accuracy, which they clearly do, but also for data quality. The paper argues that data quality, as a legal notion, should be interpreted to include data accuracy, the underlying logic of the processing technologies, e.g., organisation, formatting, abstract algorithmic models, and the application of this processing to the personal data of individuals.

The article concludes that explicitly recognising data quality as a legal concept might be only a matter of time. Recent legislative developments, including the proposed AI Act, indicate that this is not only feasible, but that quality requirements about the processing of personal data are necessary safeguards in AI technologies. The consequences of establishing data quality as a legal concept could be manifold. Undoubtedly, the controller would have to comply with the applicable data quality requirements. Furthermore, when data subjects exercise their right to rectification, the rectification measures might encompass not only the mere correction of the input data and/or the final outcome of the processing, but also of the underlying data processing technology. The paper has demonstrated how such a reading of the right to rectification is feasible, especially when read together with the accountability obligations of the controller.

From this it follows that the right to rectification could be a powerful tool to allow data subjects to rectify profiling, risk assessments and data presentation problems, in addition to factual mistakes. This could have an immense impact on business and governmental practices and on the training of humans who work with new technologies. However, it remains unclear whether and how the right to rectification applies to all the different data processing operations and the abstract criteria on profiling. Such open questions will likely be clarified through court cases.

Last but not least, for the principle of data quality to achieve its full potential, it needs to involve the data subject more actively, i.e., through the right of access to one's own data, which allows the data subject to discover and request the rectification of their data.[136]

[136] Karst (n 34), 355.